Using Dublin Core in educational material: some practical considerations based on the EASEL experience

by Aida Slavic and Clara Baiget, LITC

Access to educational material has become an important issue for many stakeholders and the focus of much research worldwide. Resource discovery in educational gateways is usually based on metadata and this is an area of important developments. Resource metadata has a central role in the management of educational material and as a result there are several important metadata standards in use in the educational domain. One of the most widely used general metadata standards for learning material is the Dublin Core Metadata Element Set. The application of this general purpose, metadata standard for complex and heterogeneous educational material is not straightforward. This paper will give an overview of some practical issues and necessary steps in deploying Dublin Core based on the LITC experience in the EASEL (Educators Access to Services in the Electronic Landscape) project.[1]

Background

The development of e-learning in recent years has created the need for efficient management, distribution, discovery and access to pedagogical materials online. New information and communication technologies have given rise to an increase in distance learning and lifelong learning which also encourages the development of e-learning and furthermore, has led to the production of educational material in a plethora of digital formats. The variety of software applications now in use for the production of educational content has contributed to the growth and diversity of this material. Apart from material that is specially created for the educational process, learning material online comprises every form of recorded communication that can be used to facilitate educational content. This means that any kind of recorded knowledge, as well as any form of its communication can ultimately become learning material. Heterogeneous in nature, learning material can be anything ranging from a single image, text, or sound, to videos or java applets. Learning resources are expensive both to create and manage and the educational community and those relying on it are looking into more efficient use of existing resources and their appropriate discovery and reuse.

Educational domain and metadata

The majority of gateways built to provide access to learning material rely upon metadata. Creation of a resource surrogate, i.e. metadata, is the only reliable way to manage resource discovery whenever the quantity or the diversity of structure and format make cheaper and simpler free text searching inefficient. Metadata has a central role in learning material management, thus several metadata standards for educational material were developed five to six years ago: IEEE Learning Object Metadata, its IMS application profile and ARIADNE metadata.[2] At the same time, some of the education community, anticipating information exchange and interoperability issues, started to use general metadata standards such as Dublin Core. Today there are several educational gateways that base their resource discovery on this general metadata standard. Some of the well known ones are the Education Network of Australia (EdNA - http://www.edna.edu.au/EdNA) and The Gateway to Educational Material (GEM - http://www.thegateway.org). The Dublin Core metadata element set used by these gateways has served as the basic core to which additional descriptive elements have been added to suit the needs of educational material.[3]

In terms of learning resource management, and as a consequence of the nature of learning material, services providing access to learning material share some common characteristics. They usually build standalone metadata repositories that are based upon the same principle as bibliographic metadata e.g. library catalogues. Standalone metadata, as implied in the name, is a description of the resources that resides and is managed separately from the resource itself. As opposed to metadata that is embedded in a resource, standalone metadata is most often adopted by third party services as it enables better control of the metadata population and allows the service to enrich and adjust the metadata to better suit its users needs.[4]

The educational domain is known for its wide application of the XML standard in formatting metadata. XML has become widely accepted as the standard in network applications and open information exchange. Platform independent, extensible, flexible and adaptable, XML encoding provides a robust basis for encoding and exchanging structured information that guarantees long life and readability of data across systems. IEEE LOM[5], IMS[6] and ARIADNE[7] standards have been created with XML encoding in mind while those who are using Dublin Core Metadata Element Set (DCMES)[8] for educational resources have to develop XML encoding to suit their needs. This process involves procedures and methods that have to anticipate information exchange objectives and the system architecture. The process of metadata implementation is basically part of a complex e-learning environment building process. It is the intention of this paper to describe some of the intricacies in implementing metadata schemas as experienced by LITC in the EASEL project. The paper will outline the steps in building a Dublin Core application profile and XML encoding for this project.

Dublin Core between interoperability and specificity

The educational domain already has very elaborate and widely accepted metadata schemas (e.g. LOM/IMS) with appropriate XML encoding, therefore one may question the reason for applying Dublin Core in learning material discovery. Information exchange and discovery on the Internet takes place in an environment where it is necessary and logical to access information irrespective of its provenance or domain. Thanks to an open network environment, each domain - no matter how self-sufficient it may well be - might appreciate the possibility of being involved in information exchange with others. This will occur whether the domain wants to broaden the scope of its resource discovery and extend it outside the local domain, or whether the goal is to make local resources widely available and accessible. The education domain is interested in all kinds of resources that might reside in libraries, museums, archives, or in some subject oriented research or scientific information service. On the other hand learning materials will be facilitated through other services and any library service supporting e-learning will have to provide access to a range of resources across boundaries of media and curatorial tradition, as recently pointed out by Pete Johnston.[9]

The Dublin Core Metadata Element Set (DCMES) has the potential to meet different requirements and is widely used across information sharing communities, including libraries, museums, archives, government, commercial information providers and education. Translated into 20 languages, it becomes a true bridge between metadata sets and communities.[10] This kind of commitment to "openness" and interoperability is one of the reasons why this schema is adopted by the educational domain.

Its advantages are said to be simplicity, semantic interoperability, international consensus, extensibility and metadata modularity, however, these advantages are also a source of problems for any metadata implementer, and the educational domain is no exception.

As it is supported by different and numerous communities, DCMES has a great appeal as a metadata standard that serves the field of resource discovery almost as a common language of understanding. But one should be cautious, bearing Tomas Baker's comparison of this metadata standard to "pigeon English" in mind, as Dublin Core tends to be hospitable to the characteristics particular to communities that deploy it.[11] As with other information exchange languages, the closer it gets to satisfying the specific needs of certain communities, the further it drifts from commonly understood descriptive elements. Although it sounds perfectly reasonable to keep the application of Dublin Core as simple as possible, this is not as easy as it may sound. The fact that the production of metadata is an expensive process often means that certain communities are not likely to engage in this process if metadata will not support their domain needs for resource discovery. This is the reason why it is often very hard for specific domains to abstain from adding complexity, new attributes, sometimes attributes to attributes, although this may represent an obstacle in processing cross-domain search indexes.[12]

The need for better structuring of data and more specific semantics was recognised very early when the broader use of DC started. Most applications required mechanisms to refine and qualify metadata elements and their values. This need initiated the development of Dublin Core qualifiers in 1997 that were supposed to meet the requirements of better semantic specificity but at the same time - because of the simple fact that more specific means less interoperable - qualifiers were provided as a separate non-obligatory addition to the 15 core elements.[13] The policy behind DC Qualifiers is that they do not change but rather refine the basic semantics of descriptive elements. In the process of information exchange this additional refinements can be stripped down leaving the basic content of each element. This rule in the application of Dublin Core is called the dumb-down principle.[14]

Considerations for using Dublin Core for educational material

Recommended qualifiers, although helpful in "deepening" the existing resource description, do not help in covering domain specific characteristics of a resource. This means that any characteristics of, for instance, educational material related to the learning process, such as learning time or specific audience to which the material is aimed, cannot be expressed in DC metadata, with or without qualifiers. To satisfy the need of specific domain communities in preserving their own community "view" of metadata while at the same time maintaining the interoperable and exchangeable parts of descriptive information, the Dublin Core Metadata Initiative (DCMI) has adopted the policy of modular extensibility.[15] Various working groups are engaged in public, cooperative and controlled development of the basic core elements, if such a need can be identified by the majority of users. The DC Education Group, for instance, was formed in order to look into educational needs and the result is a recommendation of Dublin Core extensions called DCEd. Suggested element extensions are based on the existing IEEE LOM educational metadata schema. DCEd is presently in its draft proposal stage, and suggests the following elements could be added to the basic DC core: *dc-ed.standards*, *dc-ed.audience*, *dc-ed.conformsto* (qualifier to the *dc.relation* Element). This working group also suggests the endorsement of data elements in the IEEE LOM/IMS namespace: *interactivityType, interactivityLevel*, *typicalLearningTime.*[16]

In order to function correctly, information exchange has to be supported not only by the consensus of descriptive elements, but will also depend on the semantics and interoperability of the vocabulary that is going to be used in these elements. The first step is to agree that a learning resource is going to be described from the point of view of the educational standard it complies to, or from the point of view of the audience. The next step is to provide standard vocabulary with commonly understood and widely applicable semantics. These vocabularies are still under development and many projects such as EASEL that are deploying DCMES with DCEd extensions have to cope with less standardized and less interoperable solutions. The issue of educational vocabulary is only one among many vocabulary issues that implementors of DCMES have to solve in the process of building their implementation policy.

There are few guidelines on how to create what is known as a metadata information model, although every implementor goes through this process.
The lack of fully functional and standardized supporting documentation for DCMES: a user guideline, encoding instructions and controlled vocabularies, is one of the reasons why implementing this very simple metadata model is not always a simple task.

Metadata implementation using qualified DCMES in the educational domain, and often elsewhere, is confronted with a number of issues related to the metadata implementation and its encoding in XML. These will include further semantic, syntactical and structural issues of data that is going to be recorded using DCMES and could be summarised in the following points:

- choice of elements and their refinements and additional clarification of their semantics;
- consideration of the additional descriptive elements and refinements that can be identified by specific local needs or system requirements;

- decision on whether and which elements are going to be mandatory and repeatable;
- choice of controlled vocabularies that will be used in metadata;
- choice of language, grammatical and spelling rules to be applied in metadata;
- decision on XML encoding of the elements;
- decision on metadata management.

Dealing with metadata implementation and the aforementioned questions will be based on the knowledge and requirements of the system in which metadata is going to be used, although part of it can be done independently. A resource discovery system, its purpose, its architecture and its search functions, determines the purposes and encoding of metadata. The nature and the size of the resource collection will indicate the granularity of metadata and the need for specificity. In the educational domain this is an important issue as a learning object can be a single image or single web page or even one question and its answer (e.g. in the collections of tests). Client applications and search indexes planned for the system will determine the metadata subject indexing policy. The EASEL project, which is concerned with distributed searching of learning material based upon multiple metadata schemas, did not impose many restrictions on its partners. One of the project objectives was to test whether it is technologically possible to support reliable discovery of distributed learning resources and reuse of the material in real life situations, with many content providers and versatile learning resources and different metadata schemas (see more about projects objectives and architecture at http://www.fdgroup.com/easel). The only requirement was that resource metadata be in XML format and that the XML schema necessary to parse the metadata is made available in a metadata repository. One system component, the "Course Constructor Kit" requires that learning objects submitted to the system are of very fine granularity (e.g. images, single web pages, graphs etc.) thus the accompanying metadata had to support this level of granularity in order to enable resource discovery for the course building. The EASEL service will be in English although learning material may be in other languages.

EASEL's research into current metadata developments was finalised in autumn 2000, the result of which was a deliverable that stated specifications for metadata and XML binding. It was decided that LITC would contribute metadata based on Qualified DCMES and DCEd to the EASEL repository.

The EASEL Metadata Repository has a simple standalone metadata model that supports resource discovery. It is in many ways similar to other information systems based on metadata in that it is a somewhat self-contained module. A requirement of the repository is to be a source of rich, consistent, objective and accurate information about learning material. This learning material can be from any possible field of knowledge, can come in any digital format and be of any size. It was EASEL's policy to restrict the use of metadata vocabulary to well known and widely used standards in resource description and to conform as much as possible to basic interoperability rules.

Besides the aforementioned EASEL specific requirements the rest of the procedure is probably similar to many other metadata implementations that consist of the following work fields:

- Building an information model
- XML encoding
- Indexing policy
- Metadata management
- Metadata population guidelines

The work will, obviously, first include an in depth analysis of the metadata standard and its suitability for resources that are supposed to be described with it. It is necessary for this to be recorded in detail, as with any other data information model.

Building an information model

The first and most important step in the implementation of a metadata schema is building an information model. This model contains the choice of elements that are going to be used in metadata schema. Elements can come from one or more metadata standards or can be introduced locally (very often some kind of administrative metadata). In EASEL, for instance, we are using a total of 20 elements included 15 DC elements, 2 elements from DCEd and three elements from LOM.

Also there is no reason why implementors would use all the elements that exist in one metadata standard and the information model has to represent this decision. When assembled as a document, the information model will record the way elements are going to be used, the form of their names as used in the system, data type and most importantly whether they are optional or mandatory and whether they can occur once or are repeatable. The DCMES information model also states which qualifiers are going to be used as well as what vocabulary is going to be used with each.

In the information model an implementor will decide and record whether there is a specific interpretation of the semantics of certain elements and if there is a need for clarification. For DCMES this kind of information is very often needed. For example, in the very first element: *title*, the EASEL LITC implementation of Qualified DCMES has recorded several important decisions. First, it was decided that *title* has to be mandatory and multiple which means that every metadata **has** to contain at least one *title* but can contain several. As the EASEL service is in English it was decided that the title of learning material in other languages, besides its original language, will also have the title in English. The attribute *language* is necessary for the system to distinguish, display and search for English titles. With each descriptive element, a proper information model will provide data types (e.g. string, coded list) and examples for each field.

EASEL has produced two application profiles of Qualified DCMES with DCEd: one for general use, the other is an application profile used by LITC at South Bank University.[17]

XML encoding

Once the information model was created this provided enough information to engage in XML encoding which expresses the metadata schema in terms of resource properties and their values. In order to be able to produce valid XML documents one has to create a DTD or XML Schema that will serve to parse/validate xml documents in the system. Generally an XML Schema is considered to be more flexible than a DTD and easer to use in many XML applications.[18] EASEL partners, for instance, have agreed to use the XML Schema and most of them used XML Authority to produce this basic requirement for their metadata schemas binding. Once created, the XML schema can be used to edit valid XML documents in different XML editors but more importantly it is a vital file for the system that has to process XML.

Lacking DTD examples, schemas or even instructions for the XML encoding syntax of qualified DCMES, EASEL partners have developed their own XML schema. This is based on a common sense of how DCMES should be encoded in XML to support both project purposes and interoperability principles. The main issue in encoding DCMES is to decide how to encode refinements and qualifiers as XML allows one to treat them in different ways. Related to this issue, the main concern was a dumb-down principle and the fact that any additional specification of Dublin Core i.e. sub-element or even attribute could be lost in the process of information exchange when a dumbing-down principle is used.

<table>
<tr><th>Source</th><th>Name</th><th>Explanation</th><th>Cardinality</th><th>Data Type</th><th>Notes</th><th>Example</th></tr>
<tr><td>DCMES</td><td>title</td><td>A name given to the resource.</td><td>Multiple Mandatory</td><td>String</td><td>EASEL guidelines: language of the titles should be indicated</td><td rowspan="2"><title lang=”en-GB”>Hypertext Markup Language</title>

<title alternative=”HTML” lang=”en-GB”>Hypertext Markup Language</title></td></tr>
<tr><td>DCMES-qualifier</td><td>alternative</td><td>Any form of the title used as a substitute or alternative to the formal title of the resource.</td><td>Multiple Optional</td><td>String</td><td>EASEL guidelines: alternative title will be used for acronyms only</td></tr>
</table>

Figure 1 – Layout of the EASEL DCMES information model

In Qualified Dublin Core each element describes the property of the resource and each of these properties can have further properties. Thus DC element *date,* for example can have properties: created, issued, valid, available, modified. Also one may express the value of the each property in terms of encoding scheme or controlled vocabulary

<type scheme=DCMIType>image</type>

In this example the value is the term "image" that comes from the list of resource types called DCMI Type Vocabulary. Also, DC Subject describes a property of the resource and the value used to do so can come from one of many encoding schemes thus:

<subject scheme="UDC">004.7</subject>

Those familiar with DCMES can think of many situations in which an XML nesting structure could be used with ease to encode refinements and refinements of refinements. Sometimes this will involve recommended qualifiers (marked in bold in the examples below) and sometimes this will come as a consequence of structuring data in order to preserve semantics that are locally recognised to be relevant for resource discovery (such as making a distinction between titles in different languages or making a distinction between corporate and individual creators). Some typical examples of this complexity are:

TITLE
- subtitles (s)
- language
- ***alternative title***

SUBJECT
- ***keywords***
  - language
- ***encoding scheme***

CREATOR
- individual
  - surname
  - forename
- corporate

FORMAT
- **extent**
  - **size**
  - **duration**
    - encoding scheme
- **medium**
  - **encoding scheme**

[N.B. Only properties marked in bold are the official DC recommendation, a normal type font indicates the locally specified properties]

When creating metadata one should think about the possibility that the same data may be used outside of the system as well and should preserve as much information about the learning material as possible. Lacking different, more XML oriented guidelines, the EASEL interpretation of these rules resulted in encoding Qualified Dublin Core in a flat, unstructured format. So instead of for instance:

```
<creator>
        <individual>
                <surname>Smith</surname>
                <forename>John</forename>
        </individual>
</creator>
```

EASEL has chosen the following

<creator type="individual" surname="Smith" forename="John">Smith, John</creator>

In the potential information exchange of EASEL Dublin Core metadata in an open environment, it is easy to imagine that an external client application may not want to use qualified Dublin Core or some EASEL specific attributes in which case these may be ignored without the element content being lost.

A better approach could be expected with recent developments in RDF/XML[19] encoding and in particular XML namespace[20] developments with the introduction of a namespace policy for Dublin Core[21] will allow DCMES encoding to support nesting and add more to its consistency.
However the policy of keeping a data model simple whenever this does not affect resource discovery will have to stay part of DCMES applications. This means that any additional refinements, apart from those recognised by DCMES Qualifiers, should be kept to a minimum although it is easy to imagine that each application will have a different interpretation of the absolute minimum. EASEL, for example, has decided to introduce two types of creator: individual and corporate but at the same time did not think it necessary to further encode title to title and subtitle, although the difference is preserved in the textual representation of data using punctuation:

<title alternative="HTML">Hypertext Markup Language: introductory course</title>

It should be noted that XML encoding is still evolving and a review of current developments is a necessary part of the metadata implementation. One part of the decision making process could be supported by information about DC application profiles available in metadata registries such as DESIRE[22] or SCHEMAS[23] (in development) but also by collecting more information about projects deploying Dublin Core, available at the MMI-DC Observatory.[24]

Indexing policy and metadata implementation

The description of the subject of a resource is a very important part of the metadata as this is the very core of resource discovery. Regardless of whether one refers to it as a subject, aboutness, topic, theme or discipline, this segment of metadata is the most expensive to produce as it relies on content analysis and indexing of the intellectual content of the resource, often using complex and expensive indexing tools. Deciding about objectives, scope, tools and techniques to produce this part of metadata for a particular system is an important part of the metadata implementation that is going to represent an information system's **indexing policy.** This approach is more suitable for traditional information systems than for distributed resource discovery but the EASEL experience shows that it can contribute to the richness and sanity of metadata records. It is rightfully called policy as it has to anticipate and relate system requirements, metadata formatting provision, the nature of the resources and users' needs. Indexing policy that needs to be established for this task entails not only the choice of indexing tool i.e. language(s), but also the coverage of indexing terms assigned per document (exhaustivity) and the precision of indexing terms (specificity). When some of the decisions on indexing policy are recorded this can help in the testing and evaluation of resource discovery within the system but can also help in improving metadata if such a need occurs.

From the point of view of how metadata standards support formatting of this kind of information, DCMES provides two element that are concerned with the intellectual content of the resource: element *subject* and element *coverage*. While the semantics of the first element are quite straight foward (it answers the question "what is this information object about?"), the semantics of the element *coverage* are somewhat blurred as it overlaps with *subject*. Element *coverage* is supposed to express the subject of the document when this has some temporal or spatial value.

How the subject is going to be described is only vaguely suggested by the DC standard metadata schema and its refinement. DCMES offers the possibility of using keywords and some of the widely used precoordinated indexing languages such as the Library of Congress Subject Headings, Dewey Decimal Classification, Universal Decimal Classification are suggested for use.

In the EASEL examples, LITC has put forward several arguments that are important for this decision making process. First, the subject may be considered to be the most important source of information in learning material discovery. Information in this element has to give a precise and contextualised explanation of "what the learning object is about". The consistency (always using the same term for the same content), specificity (how specific the indexing term is), and exhaustivity (how many indexing terms are used) will determine the precision and recall i.e. the efficiency of the retrieval.

Indexing languages that can be used as a source of vocabulary can be put in two categories: alphabetical indexing languages (those that use natural language terms i.e. words) and classifications (those that use symbols):

- Alphabetical indexing languages such as thesauri and subject heading systems (such as Library of Congress Subject Headings - LCSH) - are important as the user can easily understand the meaning of the term.

- Classification systems are not understandable for end users but can serve to establish terminology control, hierarchical and associative relationships between concepts. They are not language dependent so they can serve as a term control mechanism in multilingual systems

Requirements in relations for the indexing languages put forward by EASEL could be summarised as follows:

- cover the whole of knowledge (as learning objects can be about anything)

- enable very specific content description

- not be very time consuming / expensive
- easy to apply
- easy to control
- support browsing taxonomy (i.e. expressing hierarchical relationships between subjects: dynamics is part of *physics*, *physics* is part of *pure sciences*) .

It was agreed that the EASEL search will be performed by simple matching of sought terms to the subject description of the resource - this means that the subject description has to be expressed in this simple manner and not via subject headings systems and thesauri that are very complex and expensive to implement and use. It was also planned for EASEL to offer some kind of browsing taxonomy. Based on these facts it was decided to use the combination of two indexing languages:

- keywords (intuitive but as controlled as possible)
- a classification scheme (one which is specific enough, available for free in electronic form and generally applicable)

'Keyword assignment' is not controlled by any indexing system rules or vocabulary and usually produces very poor indexing. In order to have a basic source of keywords and also to impose control over keyword assignment, it was decided that UDC classification be used. The classification scheme in electronic form (edition dated 2000) that contains 62000 terms can be obtained free of charge from the UDC Consortium to be used for research purposes only. The reason UDC was chosen, apart from being easily accessible in an electronic form and as a web version are: a) size of vocabulary b) hierarchical structure c) synthetic feature. During the indexing process and metadata population the UDC schema is used not only for indexing, but also as a source of vocabulary and as a reference tool for related subject areas, synonyms or hierarchical relationships. The basic set of keywords are always extracted from the classification schedules and more specific terms are added where appropriate. This is a simple, primitive but efficient way of imposing some kind of control during indexing.[24]

Applying classification on EASEL's level is not expected to require any special skills but both keywords and the classification would require detailed guidelines for their application.

Providing guidelines for metadata population

Besides a detailed information model and the control imposed by XML encoding there are still many practical issues that have to be decided upon. The semantics of DCMES are open to interpretation and even when there is a vocabulary available the meaning and the application scope of the terms used often need further explanation. There are several areas where there will be a need for additional guidelines when implementing Dublin Core:

1. Clarifying semantics

element and refinement semantics (clarification and examples)

vocabulary semantics (clarification and examples)

2. Additional content structuring e.g.

-individual [Surname, Forename] and corporate names (corporate hierarchy),

- names as the subject of the resource (e.g. Albert Einstein or Einstein, Albert, or Einstein]

-title, subtitle (e.g. the punctuation/ style rule for expressing subtitle(s) as this is not part of the metadata encoding structure)

- style and scope of the description (what information it has to provide)

3. Text entry guidelines

language (variants of the names, UK English or American, availability of certain content in more than one language etc.)

spelling

special characters

punctuation

style (capital letters in the title, proper name in the keywords e.g. Java or java)

4. Indexing guidelines

As pointed out in the section on indexing policy, to ensure consistency of the data it is always useful to prepare guidelines. Apart from general instructions on specificity and hierarchy (e.g. should every information object on *waves*, contain also terms *mechanics* and *physics*) it will be necessary to have a general rule about numbers of indexing terms.

With regards to keywords, although this does not have to result in complete term control, indexing guidelines would contain instructions on how to deal with:

- singular and plural forms (e.g. *database* or *databases*)
- grammatical forms [nouns, noun phrases, adjectives articles]
- synonyms (e.g.*Structured Query Language* or *Standard Query Language*)
- homonyms, homographs (e.g. *reading* and *Reading; cells*-biology and *cells*- electric)
- composed terms (e.g. *distributed database management system*)
- acronyms
- spelling, transliteration, romanization
- punctuation, capitalization

In relation to classification systems, it is obvious that the system itself will already have some kind of control mechanism. However as the classification system chosen for EASEL is a synthetic system (UDC) indexing guidelines will have to indicate how specific indexing terms should be as this will indicate the amount of synthesis classification numbers.

Guidelines on semantics and additional content structuring can be easily built in to the metadata information model. Text entry guidelines could be also built into the information model by the proper and richer choice of examples that will cover text entry guidelines as well. Guidelines for subject indexing tends to be somewhat longer as they are related to the specific rules of an indexing language and they usually have to be provided separately.

Metadata management

As it was previously explained standalone metadata function similar to other information retrieval systems. In the system, metadata is kept separately from the resources, and the metadata population can also be kept detached from the system and service. In the case of XML encoded metadata, as it is possible to produce XML metadata using any of the commercially available editors (e.g. XML Instance) and then submit each XML record to the metadata repository, one can wonder about the best way of organising metadata management. Every XML editor will validate XML documents using submitted schema, but there will be no control over the textual content of the metadata elements. Therefore there will usually be no easy way of avoiding typing errors or performing an overall check on consistency of controlled vocabularies used in metadata or to change the XML metadata repository afterwards, without having special tools or engaging in additional programming.

In order to facilitate metadata population and enable better control of the metadata content consistency, LITC has created a database called EASEL DCEd Manager. This is an Access database, with a Visual Basic interface designed to manage the EASEL DCMES information model and export records in XML, based on the existing XML schema (see more details at http://litc.sbu.ac.uk/easel/dcmanager.htm).

The biggest advantage of using a relational database for metadata population is the possibility to select terms from the controlled vocabularies rather than type data entry. It is also possible to maintain vocabulary by expanding and changing existing term lists. As shown in Figure 2 one can simply choose any of the term list available from the Data Maintenance menu in order to expand terminology. In the process of the metadata population these terms are selected from the dropdown menu attached to the appropriate field.

Metadata management supported by a relational database is probably the most sensible way of handling metadata in a controlled and consistent way. It makes the process of metadata "sanitization" easier in case this becomes an issue. The

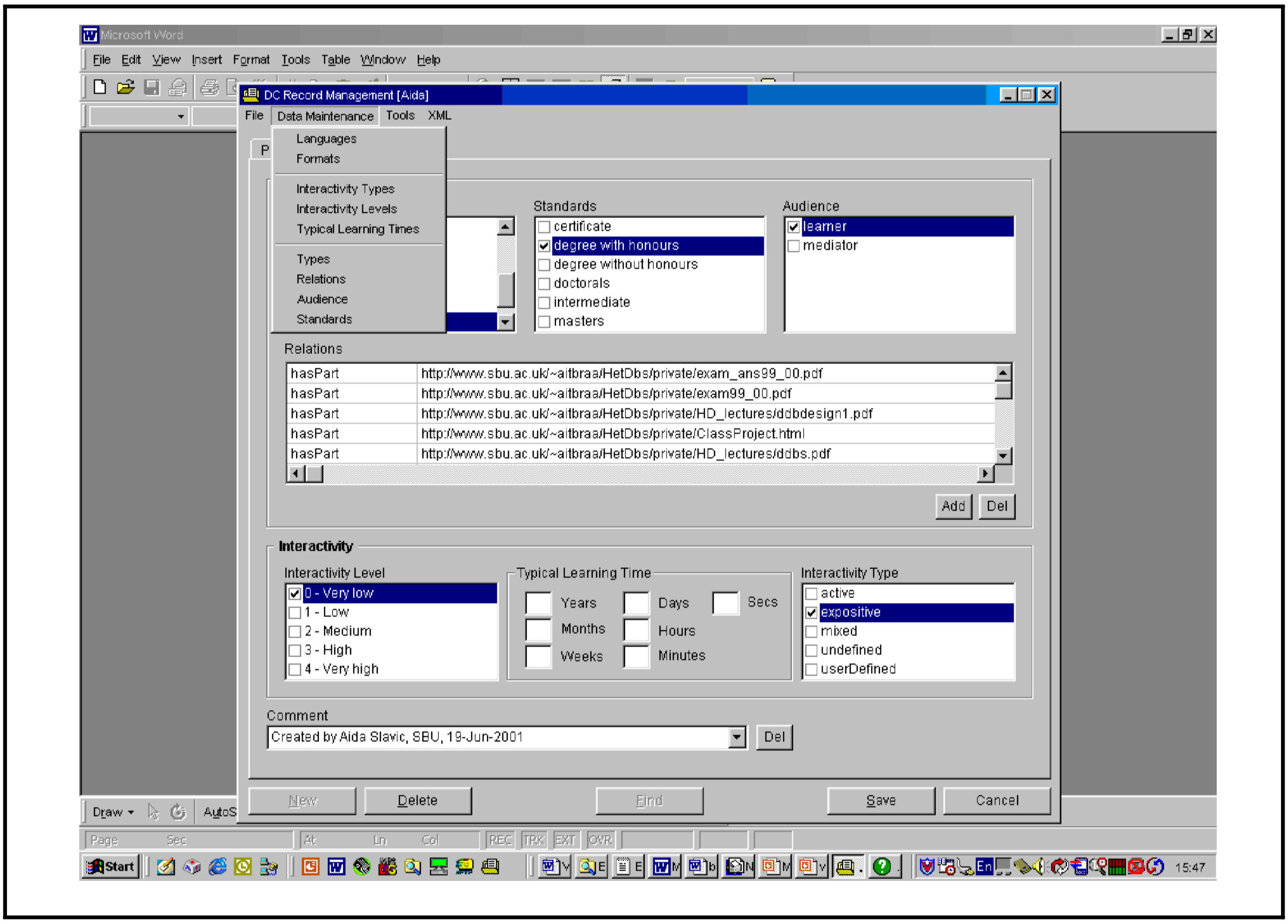

Figure 2 – Screen capture of the DCEd Manager showing Data maintenance menu

availability of a search facility and different printouts is very often needed in consistency checking. The export of records held in a database can be adjusted to any present or future system application or XML encoding.

Conclusion

Resource discovery based on metadata is the most efficient approach to resource discovery. To base a search on metadata that is kept separately (standalone metadata) from resources is often the only way to handle access and resource discovery in collections that contain information objects in a variety of digital formats. The educational domain is concerned with accurate and efficient resource discovery across heterogeneous learning material on-line and the use of metadata has been a well established approach in resource discovery while several domain specific metadata standards have been developed to meet these needs. At the same time the educational domain is interested in the application of general and more interoperable metadata standards such as Dublin Core.

Metadata production and management is very often considered to be expensive and time consuming and the use of the Dublin Core Metadata Element Set might be considered a cheaper solution than using structured and complex learning object metadata such as LOM. Using the example of the implementation of Qualified Dublin Core within the EASEL project, it is obvious that the deployment of this "simple" resource discovery metadata can present many dilemmas and may need more attention than was previously anticipated. Based on the experience of EASEL, this paper has identified several necessary steps that have to be based on both metadata expertise and system requirements: building an information model, defining XML encoding, definition of the indexing policy, organizing metadata management and providing guidelines for metadata population. The paper briefly sketched implementation steps based on the EASEL experience and could also contribute to a better understanding of the expertise needed and costs involved in DCMES metadata implementation that might be valuable outside the limited scope of the educational domain.

References

1. EASEL - Educators' Access to Services in the Electronic Netscape. <http://www.fdgroup.com/easel/>.

2. Bacsich, Paul et al. "The standards fora for online education." *D-Lib Magazine*, 5.12 (Dec 1999). <http://www.dlib.org/dlib/december99/12miller.html>.

3. Sutton, Stuart. "Conceptual design and deployment of a metadata framework for educational resources on the Internet." *Journal of the American Society for Information Science*. 50.13 (1999) 1182-1192.

4. Vellucci, S. L. "Metadata and authority control." *LRTS*. 44.1 (1999) 33-43.

5. IEEE P1484.12 Learning Object Metadata Working Group. *Draft Standard for Learning Object Metadata*. (5 February 2000). <http://ltsc.ieee.org/wg12/index.html>.

6. IMS Learning Resource Meta-data Information Model: version 1.1 - Final Specifications. <http://www.imsproject.org/metadata/mdinfov1p1.html.>.

7. ARIADNE Educational Metadata Recommendation: version 3.0: December 1999. <http://ariadne.unil.ch/Metadata/ariadne_metadata_v3final1.htm>.

8. Dublin Core Metadata Element Set, version 1.1: reference description. <http://dublincore.org/documents/dces/>.

9. Johnston, P. "After the Big Bang: forces of change and e-Learning." *Ariadne*. 27 (2001). <http://www.ariadne.ac.uk/issue27/johnston/>.

10. Bearman, D. et al. "A common model to support interoperable metadata: progress report on reconciling metadata requirements from the Dublin Core and INDECS/DOI Communities." *D-Lib Magazine*. 5.1(1999). <http://www.dlib.org/dlib/january99/bearman//01bearman.html>.

11. Baker, T. "A Grammar of Dublin Core." *D-Lib Magazine*. 6.10 (2000). <http://www.dlib.org/dlib/october00/baker/10baker.html>.

12. Lagoze, C. "Keeping Dublin Core simple: cross-domain discovery or resource description?" *D-Lib Magazine*. 7.1 (2001). <http://www.dlib.org/dlib/january01/lagoze/01lagoze.html>.

13. Dublin Core Qualifiers. <http://dublincore.org/documents/dcmes-qualifiers>.

14. Guidance information for the use of Dublin Core in Europe. CEN/ISSS Workshop Metadata for Multimedia Information – Dublic Core, Draft September 2000.

15. Dublin Core Metadata Initiative (DCMI). <http://www.dublincore.org/>.

16. Mason, J.; Sutton, S. Education Working Group: draft proposal (Revised). <http://www.dublincore.org/documents/2000/10/05/education-namespace/>.

17. EASEL DCMES information models can be seen at <http://litc.sbu.ac.uk/easel/>.

18. Powell, A. Dublin Core in XML: guidelines for implementors. 3rd draft. <http://www.ukoln.ac.uk/metadata/resources/dc/dc-xml-guidelines/>.

19. Resource Description Framework (RDF). <http://www.w3.org/RDF/>.

20. Namespaces in XML. World Wide Web Consortium 14-January-1999. <http://www.w3.org/TR/1999/REC-xml-names-19990114/>.

21. Powell, A. et al. Namespace Policy for the Dublin Core Metadata Intitiative. (Sept 2000). <http://dublincore.org/documents/2001/09/17/dcmi-namespace/>.

22. DESIRE Registry. <http://desire.ukoln.ac.uk/registry/index.php3>.

23. The SCHEMAS Registry. <http://www.schemas-forum.org/registry/>.

24. MMI-DC Metadata Observatory. <http://www.sub.uni-goettingen.de/ssgfi/observatory/>.

25. Slavic, A. "General library classification in Learning Material Metadata: the application in IMS/LOM and DCMES metadata schemas." IFLA Pre-Conference: Subject Retrieval in a Network Environment, Dublin, Ohio 14-16 August 2001 [proceedings to be published in the IFLA UBICIM series by K.G. Saur].

Contact Details

Aida Slavic
Researcher
LITC, South Bank University
Tel. 0207 815 7845
Email: slavica@sbu.ac.uk

Clara Baiget
Researcher
LITC, South Bank University
Tel. 0207 815 7843
Email: baigetc@sbu.ac.uk